IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 January 2023, accepted 8 March 2023, date of publication 17 March 2023, date of current version 29 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3258439

APPLIED RESEARCH

Accurate Detection and Precision Spraying of Corn and Weeds Using the Improved YOLOv5 Model

BAOJU WANG, YU YAN, YUBIN LAN, MENG WANG, AND ZHIHAO BIAN
School of Agricultural Engineering and Food Science, Shandong University of Technology, Shandong 255049, China
Sub-Center of National Center for International Collaboration Research on Precision Agricultural Aviation Pesticides Spraying Technology, Shandong University of Technology, Shandong 255049, China

Corresponding author: Yubin Lan (ylan@sdut.edu.cn)

This work was supported in part by the Special Funding Project for ''One Event and One Discussion'' for Importing Top Talents in Shandong Province under Grant Lu Zheng Ban Zi [2018]27, and in part by the Zibo Unmanned Farm Research Institute Project under Grant 2019ZBXC200.

ABSTRACT The precise identification of corn and weeds plays an important role in precise spraying. This paper proposed a lightweight model based on the improved yolov5s and built a precision spraying robot. Firstly, we used a data augmentation method based on category balance and agronomic characteristics to solve the data imbalance problem. Then, compared with yolov5s, yolov5l, yolov5m, and yolov5x, we found that yolov5s has both real-time and accuracy and is easier to deploy the model on edge devices. Through the feature map visualization experiment, we found that the feature extraction network can't pay close attention to the important feature of the target and suppress the feature of the noise. Therefore, we added the attention mechanism. In order to improve the real-time performance of the model, we designed the C3-Ghost-bottleneck module. Finally, we built a precision spraying robot. Compared with the original model, the value of map@0.5 is increased by 3.2%, the model file is reduced by 3.6 MB, the AP value for corn is increased from 93.2% to 96.3%, and the AP value for weeds is increased from 85.6% to 88.9%. Finally, the precision spraying experiment of weeds was carried out. The recognition accuracy of weeds is 83%, the probability of the spraying robot correctly identifying weeds and accurately spraying is 81%, and the detection speed is 30ms/f. The experimental results verify the feasibility of precision spraying weeding and the effectiveness of the improved model, which can provide a reference for the engineering application of precision weeding.

INDEX TERMS Data balance, object detection, precision spraying robot, SENet, yolov5s.

I. INTRODUCTION

In China, corn is one of the most significant food crops. Its planting area makes up 33.6% of the entire area used for food crops, while its production makes up 36.1% of the total grain yield [1]. Field weeds are one of the common disasters in corn production. They make corn less productive by competing with it for light, water, and minerals in the soil [2]. At present, the weeding methods for corn mainly include chemical pesticide weeding and physical weeding. Manual weeding requires a lot of labor, increasing costs, and low weeding efficiency; Although mechanical weeding is more effective, it cannot effectively remove the weeds between the plants, which makes it easy to harm crops. Spraying weeding operations mostly adopts continuous spraying methods. Although it has a good weeding effect, a large number of spraying pesticides not only pollute the environment but also affect the growth and development of crops, and is easy to produce pesticide residues [3].Precision spraying is an effective way to improve the utilization rate of pesticides, and the accurate identification and location of the target is the premise to achieve precision pesticide application technology. At the moment, machine learning [4], [5], [6], [7], ultrasonic sensor [8], [9], [10], laser radar [11], spectrum

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

analysis [12], [13], [14], and spectral imaging [15], [16], [17] are the most widely utilized target accurate identification and location technologies. Spectral analysis technology is susceptible to outdoor light and has the disadvantages of a large amount of spatial data and difficult analysis. Machine vision technology has been widely used in the identification of crops and weeds due to its low cost and complete information [6], [18].

In the early stage, many scholars at home and abroad used the shape features, texture features, color features, and location space features of plant leaves to identify crops and weeds. Strand et al. [19] classified weeds using a Bayesian classifier based on characteristics of the plant's morphology, and their success rate was 76.5%. Li et al. [20] used the color features of plants and land to segment them, and then used the area features of crops and weeds to classify them by constructing a pixel histogram. The experimental results show that the recognition rate of corn is higher than 95 %. Lanlan et al. [21] used four shape parameters of corn and weeds to realize the recognition of corn and weeds by SVM, and the recognition accuracy reached 96.5%. Wu et al. [22] used texture features to accurately identify corn and weeds, and SVM classifier identification accuracy is between 92.31% -100%. Due to the similarity in color between corn and weeds, the ease with which corn leaves can obscure the weeds' leaves, and the complexity of the background, it is challenging to distinguish between crops and weeds using a single attribute. Therefore, a multi-feature fusion strategy for target recognition has been suggested by several researchers. Mao et al. [23] achieved the recognition of crops and weeds using color, location, texture, and form features. The spectral imager method was utilized by Lin et al. [24] to merge the texture and form data of the leaves, and the model's recognition accuracy reached 95%. Chen et al. [25] used the Otsu threshold method to distinguish plants from the background. According to the leaf shape of corn and weeds, a probabilistic neural network method was used to distinguish corn and weeds. The recognition rates of corn and weeds were 92.5 % and 95 %, respectively.

Although the traditional machine learning method can also have a good recognition effect, due to the complex field environment, a wide variety of weeds, and the light intensity changing at any time, the traditional method lacks strong feature extraction and generalization ability, resulting in poor adaptability of traditional methods to changes in environment and weed species. Compared with traditional machine learning, deep learning uses a convolutional neural network to extract multi-scale and multi-dimensional spatial semantic feature information of weeds and independently obtains useful features of the target, which solves the disadvantages of traditional methods to extract weeds and crop features and effectively improves the recognition and detection accuracy of crops and weeds. In recent years, deep learning methods have been widely used in the identification and location of crops and weeds [26]. Jiang et al. [27] used a graph convolutional neural network to identify corn and weeds, and the recognition accuracy reached 97.8%. Andrea et al. [28] used cNET to distinguish corn and weeds in real time and used the data set generated in the segmentation stage to train the convolutional neural network to realize the recognition of corn and weeds. Pei et al. [29] built an intelligent weeding robot system based on the yolov4 network model. When the robot moving platform moves forward at a speed of 1.2 km /h, the recognition rates of corn seedlings and weeds are 96.04% and 92.57%, respectively. Cheng et al. [30] proposed an improved YOLOv4 model by replacing the CSPDarknet53 network with MobileNetv3 and then using transfer learning to accelerate model training. Experiments show that the average recognition accuracy of corn and weeds is 89.98 %, and the detection speed is 69.76 f/s. Quan et al. [31]used VGG19 as a pre-training network based on the Faster-R-CNN network model, and the accuracy of recognizing corn seedlings reached 97.71%.

In this paper, corn at the 3-5 leaf stage and weeds growing simultaneously with corn were used as identification targets. This paper presents an improved target detection model for yolov5s. The SENet attention mechanism was introduced at the feature extraction stage to direct the model to pay attention to target information, improving the model's performance for target recognition. By utilizing the lightweight benefit of the Ghost module, which decreases the model's parameters and calculations, the C3-Ghost-Bottleneck module was created to address the issue of a large number of model parameters and calculations. The trained model was embedded into NVIDIA-AGX, and an intelligent spraying robot was built. This study provides a reference for precision weeding.

II. MATERIALS AND METHODS

A. IMAGE AND DATA ACQUISITION

The experimental collection site is located in an ecological unmanned farm of the Shandong University of Technology(36°58'48''N, 118°15'36'' E) in Linzi District, Zibo City, Shandong Province, China.The geographical location of the image acquisition is shown in Figure 1.The sampling device in this study is a high-resolutionSony A7 camera with a Sony full-frame standard zoom lens. the exposure parameter is automatic,and the objective focus system is set to autofocus mode, the shutter speed is automatic mode,and the ISO parameter is automatic mode; The researchers simulated the image acquisition module of the spraying robot, and the handheld camera continuously changed the shooting angle andshooting distance to obtain corn and weed data under different environmental conditions.

The best wedding period for corn is the three-to-five-leaf stage, therefore, according to the characteristics of the weeding period of corn, we collected images of corn at the three-to-five-leaf stage and weed images during the synchronous growth period of corn, because the first weeding has been done before the corn emerges, so the weeds at this time are mostly at the seedling stage, the weeds species we collected were horse common crabgrass, and green

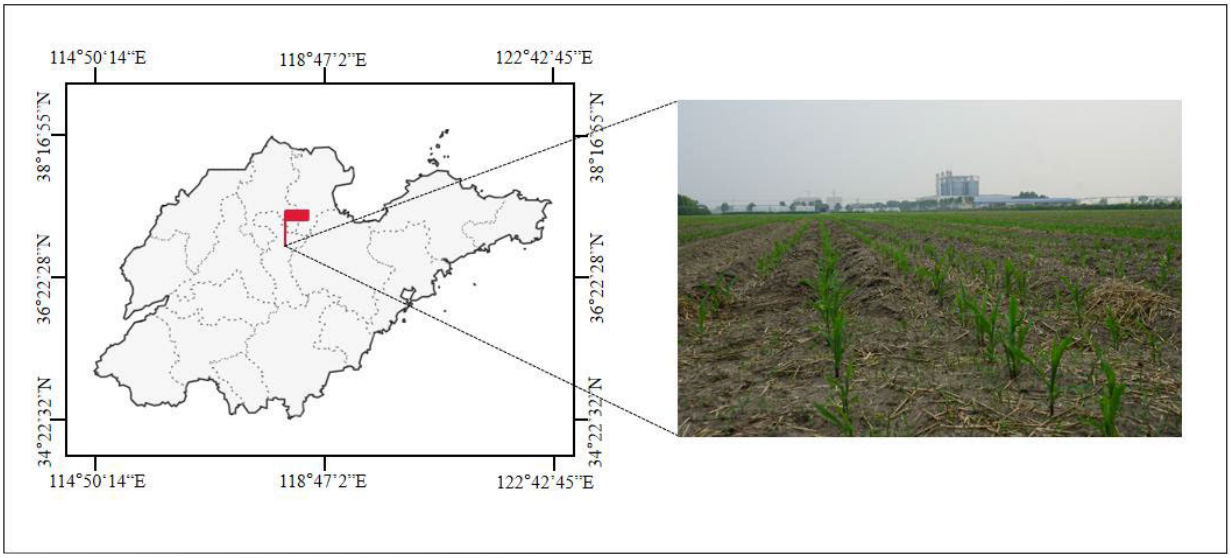

FIGURE 1. Location of images acquisition site.

TABLE 1. The time of image acquisition.

time	reason
6:00-7:00 and 17:00-18:00	Low light, obtained data when light is not good
9:00-11:00 and 15:00-17:00	Plenty of light, obtained excellent data
12:00-14:00	Obtained data for shadow occlusion

bristlegrass. The photographs were taken on July 6-10, 2020, and July 8-12, 2021, images were acquired at different times of the day and under sunny, cloudy, and drizzly weather conditions, respectively, so we obtained RGB images with different light intensities, colors, poses, sizes, backgrounds, densities, etc. The detailed time of image acquisition is shown in Table 1. 15,000 images of different scenes were collected, and 13,000 images were left after data screening.

B. DATA ENHANCEMENT METHODS BASED ON CATEGORY BALANCE AND AGRONOMIC CHARACTERISTICS

According to the number of corn and weeds in 13,000 labeled images, there are 9,000 corn and 4,000 weeds labels. The proportion between the two is close to 2.25:1. It can be found that there is a larger data imbalance between the number of corn and weeds. If this data is used directly for data enhancement, it will increase the imbalance between the data, resulting in a reduced ability of the model to detect weeds.

Aiming at the problem of data imbalance, this paper proposed a data balancing method based on the category to optimize the dataset, which makes the detection performance of the model improved [5]. This method needs to obtain the quantitative values of all categories first, determine the number of the dataset after enhancement, and then amplify the data for each category. The specific steps are as follows:

1) Input: Suppose there is a dataset, $D = [S_1, S_2, S_3, \ldots S_i][N_1,N_2,N_3, \ldots N_i]^T$, where $S_i$denotes the number of types of samples in the dataset, and N_i denotes the number of samples in each category;
2) Determine the total number of datasets after data enhancement, and C denotes the total number of datasets. Suppose the number of sample types is t, use C to divide by t, and then division M is obtained, the M denotes the number of samples after amplification for each category.
3) Observe the sample quantity values of all categories in the dataset S_iN_i, use M to divide by S_iN_i, and then division N is obtained, the N denotes the multiple to be amplified for each type of sample. The calculation is given in Equation (1):

$$\begin{aligned} M &= \frac{C}{\mathrm{t}}, \\ N_{\mathrm{i}} &= \frac{M}{S_i N_i}, \\ N &= [\{N_1, N_2, N_3, \ldots, N_i\}], \end{aligned} \tag{1}$$

4) Determine the types of data amplification methods as n, use N to divide by n, and then division T' is obtained, T' is a multiple of each data enhancement method.
5) Each type of sample in the dataset is enhanced in a different way so that the amount of sample data for each class reaches M.
6) Output: The final output is the expanded dataset, $D' = [S_1,S_2,S_3,\ldots S_i][N'_1,N'_2,N'_3,\ldots N'_i]^T$

To prevent overfitting or non-convergence phenomena caused by too little training data, this study performed data augmentation processing on the dataset. The environmental factors that affect the recognition effect when carrying out precision recognition operations in the field are mainly: the variation of light intensity, the obstruction of the target by field debris, and the different sizes of weeds. To avoid the image producing too many redundant features and reduce the recognition performance of the model, according to the field agronomic characteristics of corn and weeds, data enhancement mainly adopts five methods: Random rotation, random cropping, random brightness adjustment, random scaling, and random occlusion. The data enhancement by the above method expands the number of pictures to 38,000 and makes the ratio of the number of corn to weeds close to 1:1, so the number of corn is expanded 2 times to 18,000 and the number of weeds is expanded 5 times to 20,000, the proportion of mature and immature plums in the training set has changed from 2.25:1 to 0.9:1 so that the number of different categories of the dataset is similar.

C. DATASET PRODUCTION

The LabelImg, an image annotation tool, was used for manual annotation to obtain the ground truth for subsequent training. The annotation information was saved in the format of the YOLO dataset and marked as two types corn and weed. For occluded corn and weeds, only the exposed parts of the image are marked. The unmarked processing was performed when the degree of occluded weeds was more than 70%. For several corn or weeds growing together, mark the targets one by one.

For the marked 38,000 images, the dataset was divided into the training set, validation set, and test set, where the ratio of the three is 8: 1: 1, and there is no intersection between the data of the three.

D. YOLOV5 MODLE

YOLO series is one of the commonly used algorithms in the field of target detection. The yolov5 algorithm combines the features of the yolov1-yolov4 versions and has been improved in terms of detection speed and accuracy, and it is widely used in the field of target detection with its excellent performance. We divide yolov5 into four parts: input, backbone, Neck, and prediction.

The input end of yolov5 mainly resizes the input image to the size required by the network. The main operations include Mosaic data augmentation, adaptive anchor frame calculation, and adaptive image scaling, which enables the model to perform image reading and training.

The Backbone structure uses deep convolution operation to extract the target high, medium, and low-level feature information, mainly including the C3 module and SPP module. The C3 Module is the main structure for residual feature learning. The SPP module is called spatial pyramid pooling and it consists of 4 parallel branches, which are 1×1 skip connection, $5 \times 5.9 \times 9$,13×13 maximum pooling. The maximum pooling method of $k = \{1 \times 1, 5 \times 5, 9 \times 9, 13 \times 13\}$ is used, where 1×1 represents no processing, and then the feature maps of different scales are stacked.

The Neck network is mainly FPN + PAN structure. The structure is based on the feature pyramid network and adds a path aggregation network to improve the detection performance of the model for targets of different sizes. The FPN structure uses upsampling to pass down the strong semantic features of the higher layers so that the model can detect multiple different scales and enhance the entire pyramid. The PAN structure integrates the feature maps of high and low layers so that each feature map has rich semantic information and strong localization features.

The Prediction section generates the class probability and location information of the predicted targets, including three detection branches, to detect targets of different sizes. The loss function of yolov5 consists of three parts: classification loss function, regression loss function, and confidence loss function. In the post-processing process of target detection, a target will have multiple candidate marker frames, and to filter the most suitable target frame, the yolov5 model chooses the NMS method to suppress the non-maximal elements and search for local maximal values, and the specific process is as follows.

1) Firstly, the confidence scores of candidate boxes with scores below the threshold are reduced to zero.
2) And then the candidate boxes with the highest scores are retained by ranking them according to their confidence scores.
3) Traverse the remaining candidate boxes, calculate the IoU value of the candidate boxes retained in the previous operation, and remove the candidate boxes whose IoU value is greater than the threshold.
4) Repeat the above operation until all the reserved candidate boxes are found.
5) According to the reserved candidate box, mark the target box in the image.

The Yolov5 project has four models, they are yolov5s, yolov5m, yolov5l, and yolov5x. Among them, the yolov5s network model is the one with the smallest layer depth as well as width. When the yolov5 model does not change the depth and width of the network, the network model is shown in Table 2.

In table 2, - 1 in the source represents the upper layer. In the four deepening network models, the width of the model is changed by modifying the number of convolution kernels in the Conv structure, and the depth of the model is changed by changing the number of C3 structures, to realize the selection of models with different widths and depths according to project requirements.

E. THE PROPOSED ALGORITHM

1) ATTENTION MECHANISM

To obtain more detailed information about the target that needs attention and suppress other useless information from different channels, we introduced the Attention network, SE layer. SENet mainly focuses on the feature fusion among

TABLE 2. Network structure of yolo5.

layer	structure	stacking times	output image size	source	input dimension	output dimension	size of the convolution kernel
0	Focus	1	320x320	-1	3	64	3x3
1	Conv	1	160x160	-1	64	128	3x3
2	C3	3	160x160	-1	128	128	[1,3]
3	Conv	1	80x80	-1	128	256	3x3
4	C3	9	80x80	-1	256	256	[1,3]
5	Conv	1	40x40	-1	256	512	3x3
6	C3	9	40x40	-1	512	512	[1,3]
7	Conv	1	20x20	-1	512	1024	3x3
8	SPP	1	20x20	-1	512	1024	[5,9,13]
9	C3	3	20x20	-1	1024	1024	[1,3]
10	Conv	1	20x20	-1	1024	512	1x1
11	Upsample	1	40x40	-1	512	512	——
12	Concat	1	40x40	-1,6	512	1024	——
13	C3	3	40x40	-1	1024	512	[1,3]
14	Conv	1	40x40	-1	512	256	1x1
15	Upsample	1	80x80	-1	256	256	——
16	Concat	1	80x80	-1,4	256	512	——
17	C3	3	80x80	-1	512	256	[1,3]
18	Conv	1	40x40	-1	256	256	3x3
19	Concat	1	40x40	-1	256	512	——
20	C3	3	40x40	-1	512	512	[1,3]
21	Conv	1	20×20	-1	512	512	3x3
22	Concat	1	20×20	-1,10	512	1024	——
23	C3	3	20×20	-1	1024	1024	[1,3]
24	Detect	1	20,40,80	17,20,23	——	——	——

channels of the convolution operation in the backbone network. The network can automatically learn the importance of different channel characteristics by focusing on relationships between channels.

The SENet model compresses each feature map to establish dependencies between channels. The structure of the SENet model is shown in Figure 2.

Where C', H' and W' represent the number of channels, height, and width of the input feature map, X represents the input feature map, U represents the feature map obtained after a series of convolution and pooling operations, W and H represent the width and height of the feature map, C represents the number of channels of the feature map and $\widetilde{X}$represents the output feature map.

As can be seen from Figure 2.5, SENet consists of three parts: Squeeze, Excitation, and Scale.

a: SQUEEZE

The global average pooling operation is performed on the feature map of CxHxW to compress the features in the spatial dimension, and the feature map vector is changed into $1 * 1 * C$ so that it has a global receptive field. The calculation method is as follows:

$$Z_c = Fsq(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j) \tag{2}$$

where Z is generated by compression operation, $Z_c$is the cth element in Z,$u_c$is the cth two-dimensional matrix in U, and$u_c(i, j)$is the (i, j)th element in u_c.

b: EXCITATION

After the Squeeze operation obtains the channel information, it uses two fully connected layers to form a gate mechanism

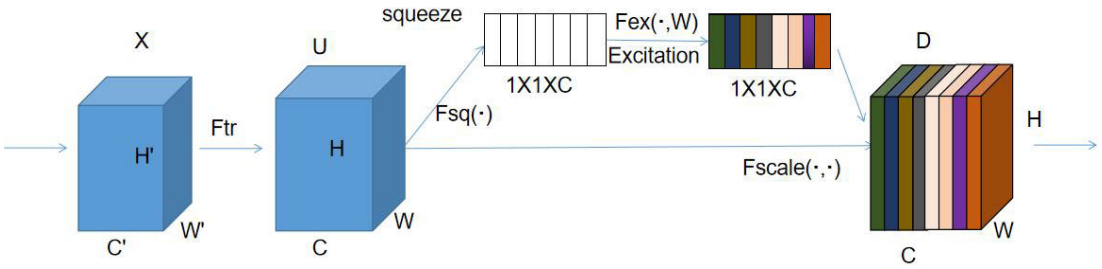

FIGURE 2. The structure of the SENet.

and activates it with Sigmod. The Excitation operation uses weight values to characterize the dependency between channels, and the weights for channels are obtained by learning through two fully connected layers. The calculation method is as follows

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2\delta(W_1 z)) \tag{3}$$

where A represents the weights of the two fully connected layers, which can limit the model complexity and assist in the role of model generalization, and δ represents the activation function.

c: SCALE

The output weights after the Excitation operation are used as the importance of each feature channel, and then the output weights are multiplied onto the input features of the corresponding channels using the multiplicative weighting method to obtain the important features and suppress the unimportant features.The calculation method is as follows:

$$\widetilde{X}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \tag{4}$$

2) IMPROVEMENT BASED ON C3 MODULE

The application environment of corn and weed identification is complex, and the hardware performance is relatively low. Therefore, under the premise of ensuring accuracy, the size of the model should be minimized, and fewer parameters should be used to generate more features to improve the calculation speed.

In the convolution operation, many feature maps are generated, and the comparison of a large number of feature maps reveals that some of them have similar features in the channel direction. However, the Ghost module uses relatively simple and less computational operations instead of convolution operations to generate feature maps, thereby accelerating the inference speed. The structure of the Ghost module is shown in Figure 3.

As seen in Figure 3, compared with the ordinary convolution operation, the Ghost structure obtains feature maps with fewer channels by convolution operation and then obtains richer feature maps by the cheap operation, and finally stitches the obtained feature maps together. The principle of reducing model computation for the ghost module is described below.

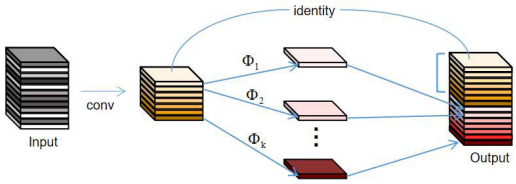

FIGURE 3. The structure of the Ghost module.

The calculation formula for FLOPs of traditional convolution is shown in (5):

$$F_{conv} = K \cdot K \cdot C\_in \cdot C\_out \cdot H \cdot W \tag{5}$$

Among them, K·K represents the size of the convolution kernel, C_in and C_out represent the number of input and output channels, and H·W represents the size of the output feature map.

In the ghost module, the feature map is first obtained by convolution operation, which has C_out/s channels, and the calculation formula for FLOPs is shown in Formula (6):

$$F_{ghost} = K \cdot K \cdot C\_in \cdot \frac{C\_out}{s} \cdot H \cdot W \tag{6}$$

Each feature map uses the convolution kernel of d·d to generate s-1 feature maps, where s is the number of linear transformations, and the calculation formula for FLOPs is shown in formula (7):

$$F_{ghost}1 = d \cdot d \cdot (s-1) \cdot \frac{C\_out}{s} \cdot H \cdot W \tag{7}$$

The compression amount of the model is quantitatively calculated, and the compression ratio is used as the calculation index. The calculation formula is shown in (8):

$$\begin{aligned} R_c &= \frac{C\_in \cdot C\_out \cdot \mathrm{K} \cdot \mathrm{K}}{C\_in \cdot \frac{C\_out}{s} \cdot \mathrm{K} \cdot \mathrm{K} + \frac{C\_out}{s} \cdot (s-1) \cdot d \cdot d} \\ &\approx \frac{s \cdot C}{s + C + 1} \approx s \end{aligned} \tag{8}$$

Finally, the acceleration ratio is used to approximately replace the computing speed. The calculation formula is shown in (9):

$$\begin{aligned} R_s &= \frac{K \cdot K \cdot C\_in \cdot C\_out \cdot H \cdot W}{C\_in \cdot H \cdot W \cdot K \cdot K + H \cdot W \cdot (s-1) \cdot d \cdot d} \\ &\approx \frac{s \cdot C\_in}{s + C\_in - 1} \approx s \end{aligned} \tag{9}$$

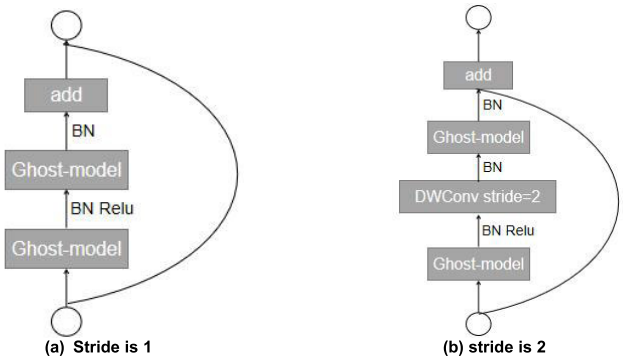

FIGURE 4. The structure diagram of the Ghost-bottleneck.

Assuming that K is approximately equal to d, and s is much smaller than C_in, through the above formula we can find that the ghost module reduces the amount of calculation, the amount of calculation is about 1/s of the traditional convolution. It shows the superiority of the Ghost module in theory. Therefore, in this paper, the Ghost-bottleneck module, referred to as Gb, is designed by using the ghost module and bottleneck structure. There are two Ghost-bottleneck structures, as shown in Figure 4. One of them is a structure with a step size of 1, which is shown in Figure 4(a). It mainly consists of two Ghost modules. The first Ghost module is mainly used as an extension layer to increase the number of channels. The main function of the second ghost module is to reduce the number of channels. The other is a structure with a step size of 2, as shown in Figure 4(b). The difference between the two structures can be seen in the graph. In Figure 4(b), a convolution is added between the two Ghost structures, which is mainly used for downsampling.

The ghost-bottleneck structure was merged with the C3 structure to design the C3-Ghost-bottleneck module, referred to as C3_Gb. The C3 module in the yolov5s model was replaced by the C3_Gb module, and its structure is shown in Figure 5.

By adding the SENet module to the feature extraction network, the model can pay more attention to important features, suppress the extraction of unimportant features, and improve the recognition performance of the model for weeds. The Focus module in the backbone was replaced by the Conv module to facilitate the embedding and exporting of the model. Finally, the C3 module was replaced by the C3_Gb module to reduce the number of parameters and calculations of the model, so that the model reduces the performance requirements of the hardware and makes the model easy to deploy on the edge. The improved yolov5s structure is shown in Figure 6.

F. PRECISION SPRAYING ROBOT CONSTRUCTION

To verify the reliability of the improved algorithm and the feasibility of precision spraying in the field, a precision spraying robot was built. The precision spraying robot is shown in Figure 7.

It primarily consists of two components: a multipurpose farm cart with four wheels and a precision spraying system. The precision spraying system is mainly divided into four parts: image acquisition unit, a target detection unit, spraying control unit, and spraying unit.

The image acquisition unit uses a USB 4K HD distortionless camera, and the video images captured by the camera are transmitted to the target detection system using USB. The improved yolov5s algorithm is embedded into the NVIDIA Jetson AGX Xavier to form a target detection unit, and the collected image video is transmitted to the target detection unit. After the model detection, the target information of corn and weeds is generated, and then the target information is transmitted to the spraying control unit. The hardware used for the spraying control unit is the Arduino UNO R3 (CH340G). The Arduino UNO R3 receives a signal from the AGX when the target detection system identifies a weed, and the Arduino UNO R3 then sends a spray signal to regulate the pump for accurate application. The spraying unit consists of a high-pressure brushless water pump, a pressure nozzle, and a 22-liter medicine box. The main function of this unit is to receive signals from the Arduino UNO R3. When a weed target is detected, the Arduino UNO R3 will immediately send a signal to the pump to control the pump to spray.

III. RESULT AND DISCUSSION

A. EVALUATION INDICATORS

In the field of target detection, the commonly used evaluation indicators are precision (P), recall rate (R), average precision (AP), and mean Average Precision (mAP). Their calculation formulas are as follows.

$$\Pr ecision = \frac{TP}{TP + FP} \tag{10}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{11}$$

$$AP = \int_0^1 P(R)\mathrm{d}(R) \tag{12}$$

$$mAP = \frac{\sum_1^n \int_0^1 P(R)dR}{n} \tag{13}$$

Among them, TP represents the number of correctly detected corn and weeds; FP represents the number of misclassified corn and weeds; FN represents the number of missed corn and weeds; AP represents the area composed of the PR curve and the coordinate axis. The higher the AP value is, the better the performance of the target detection algorithm is. The mAP represents the AP average of multiple categories, and its value represents the general detection performance of the algorithm for different categories.

B. TRAINING ENVIRONMENT AND PARAMETERS SETTINGS

For model training, the hardware and software platform configuration was as follows: CPU is AMD Ryzen Threadripper 3970×32-Core Processor 3.69Ghz, memory is 64GB, storage

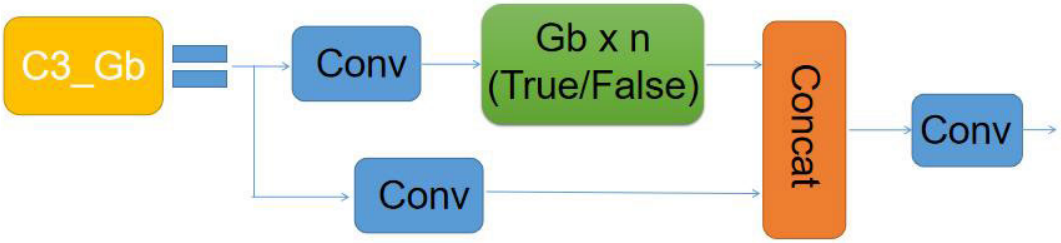

FIGURE 5. The structure diagram of C3_Gb.

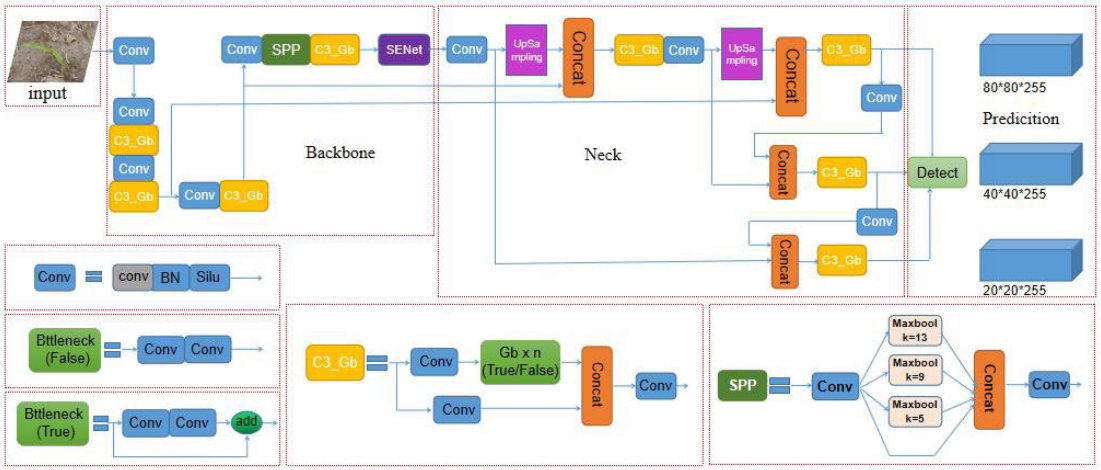

FIGURE 6. The structure diagram of improved yolov5s.

FIGURE 7. Pesticide spraying robot.

SSD is 4TB, display card is Nvidia TITAN RTX, display memory is 24GB, the operating system is windows10, the Cuda version is 10.2, the python version is 3.7, the PyTorch version is 1.10

In the experiment, the incipient learning rate was set to 0.01, the final learning rate was set to 0.2. The momentum parameter was set to 0.937. The weight decay parameter was set to 0.0005. To reduce the overfitting phenomena of the model in the beginning stage for small batches of data and to prevent model oscillation to maintain the deep stability of the model, a warmup strategy was used during training. The input image was uniformly adjusted to 640 × 640 pixels, the epoch was set to 300, and batch size was set to 64 during training.

C. yolov5 MODEL RESULT ANALYSIS

To compare the recognition performance of the four network models, yolov5s, yolov5m, yolov5l, and yolov5x for corn and weeds and choose the more appropriate algorithm to be embedded into the precision spraying robot. In this study, the model was trained based on the constructed dataset, the comparison of the results is shown in Table 3.

TABLE 3. Comparison of the structure and performance of different versions of yolov5.

	Yolov5s	Yolov5m	Yolov5l	Yolov5x
width	0.5	0.75	1.0	1.25
depth	0.33	0.67	1.0	1.33
C3-n(True)	1,3,3	2,6,6	3,9,9	4,12,12
C3-n(False)	1	2	3	4
Number of Conv convolution kernels	32, 64, 128 256, 512	48, 96, 192, 384, 768	64, 128, 256 512, 1024	80, 160, 320 640, 1280
GFLOPS	16.4	50.4	114.3	217.3
Params（M）	7.2	21.2	46.5	86.7
mAP@0.5	89.4%	93.5%	94.3%	96.5%
Speed（ms）	6	9	12	19
training generated model file (MB)	14.4	40.5	89.3	180

All models were trained on Nvidia TITAN RTX graphics cards. According to the table, the width of the model increases from 0.5 to 1.25, the depth of the model increases from 0.33 to 1.33, and the value of mAP@0.5 gradually increases from 89.4% to 96.5%. The detection performance of the model is improving, but the detection speed increases from 6ms to 19ms, and the speed is slowing down. Yolov5x was found to have the highest mAP@0.5 value in the comparison, however, the detection time for a single image is 3 times slower than that of the yolov5s model, and the training generated model file is 13 times larger than that of the yolov5s, but the mAP@0.5 value of the two models differed by only 7.1%.

Therefore, yolov5s has both real-time performance and detection performance, which is easy to deploy on the mobile terminal and edge terminal, and easy to test the landing of products. The yolov5s model was chosen as the corn and weed identification model in this paper.

D. FEATURE MAP VISUALIZATION EXPERIMENT

In the field of target detection, the goodness of feature extraction often determines the performance of target detection. Thus, we visualized the feature map in the feature extraction network of yolov5, to find the problem of the feature extraction network and provide improvement direction for the model. The characteristic diagram of corn is shown in Figure 8, and the characteristic diagram of weeds is shown in Figure 9.

As seen in Figures 8 and 9, the Focus module processes the image, and a 64-dimensional feature map is produced as a result. It can be seen from the output feature map that some data in the picture is lost, but the contours of corn and weeds can be seen. After the Focus module, it continues to go through Conv, C3-1, Conv, C3-3 and other modules, the information in the image continues to be lost and the detectability of the image becomes less and less, and the features of corn and weeds become more and more blurred, indicating that the model first extracts shallow feature information such as color and shape, and then extracts semantic features of higher level and higher dimension. and the information of the background map is gradually hidden by the network layer, and the target information to be detected is finally extracted.

From the feature map visualization experiments, it can be seen that the yolov5s model does not focus on extracting features of corn and weeds and cannot give different weights to irrelevant and important features. From Figure 8 and Figure 9, it can be seen that the model extracts the feature information of noise such as straw many times, while the feature information of weeds is not fully extracted, if the feature information extraction of irrelevant noise can be reduced, it will be more favorable for the model to extract the features of the target. In the feature information extraction process, the feature information is mainly extracted by a convolution operation. The feature information lacks the correlation between different dimensions such as channel and space, and cannot increase the weight of key feature information, with the deepening of the convolution layers, the feature information is gradually lost, which eventually leads to inadequate feature extraction of weeds by the model. Therefore, in the feature extraction network of yolov5s, the correlation of key feature information between channels should be strengthened and the extraction of unimportant features should be suppressed to enhance the recognition performance of the model for the target.

E. RESULTS ANALYSIS OF IMPROVED yolov5s MODEL

1) TRAINING RESULT ANALYSIS OF MODEL

Using the corn and weed photos from the complicated field environment as the training set, the improved yolov5s model

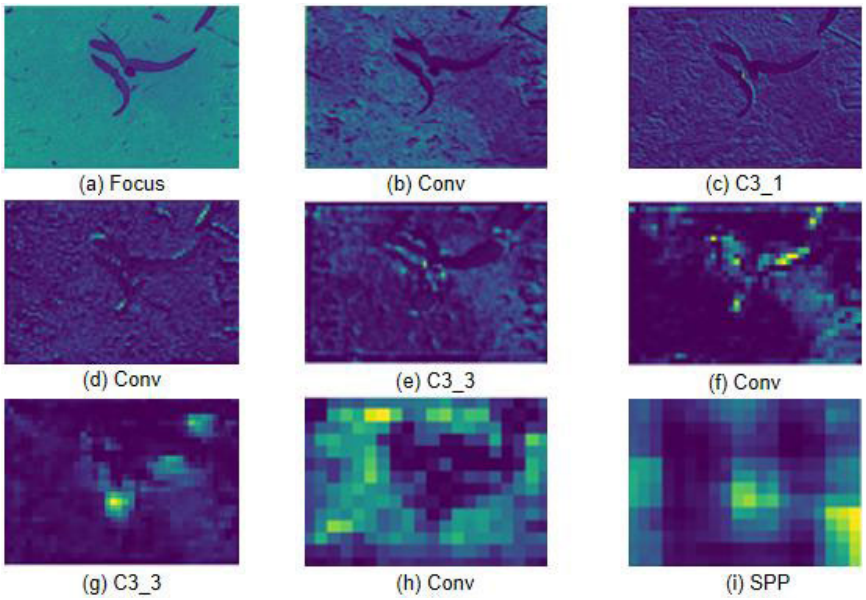

FIGURE 8. Feature map visualization of corn in the feature extraction stage.

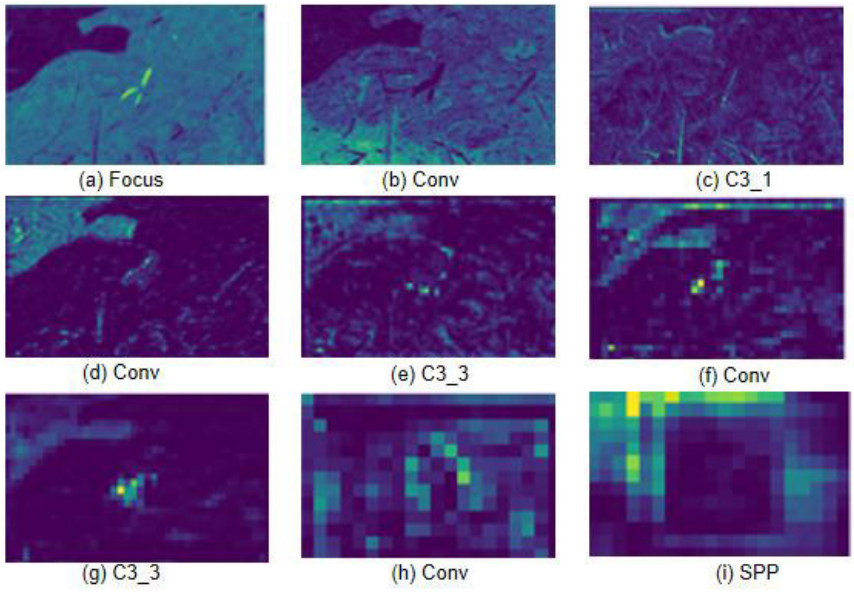

FIGURE 9. Feature map visualization of weeds in the feature extraction stage.

was compared to the yolov5s model. The superiority of the improved yolov5s model on the training set was confirmed. In Figure 10, the change in the value of mAP@0.5 through time is depicted and the improved yolov5s model's loss value curves for the training set and validation set are shown in Figure 11.

It can be seen from Figure 10 and the training log that after 160 training iterations, the value of mAP@0.5 of the improved model reaches more than 0.9. After 300 training iterations, the value of mAP@0.5 reaches 92.6%, and the value of mAP@0.5 of the yolov5s model is 89.4%. The value of mAP@0.5 of the improved model is 3.2% higher than that of the original model. The overall detection performance of the improved model is improved, indicating that the model can correctly identify corn and weeds, as well as achieve accurate identification standards for corn and weeds.

The detailed performance comparison results compared to the yolov5s model are shown in Table 4.

As can be seen from the table, the mAP@0.5 value of the improved yolov5s model reaches 92.6%, which is 3.2% higher than the original model. The AP value for corn recognition increases from 93.2% to 96.3%, and its AP value increases by 3.1%. The AP value for weeds increases from 85.6% to 88.9% and its AP value increases by 3.3 %. It can be seen from the above data that the mAP@0.5 value of the

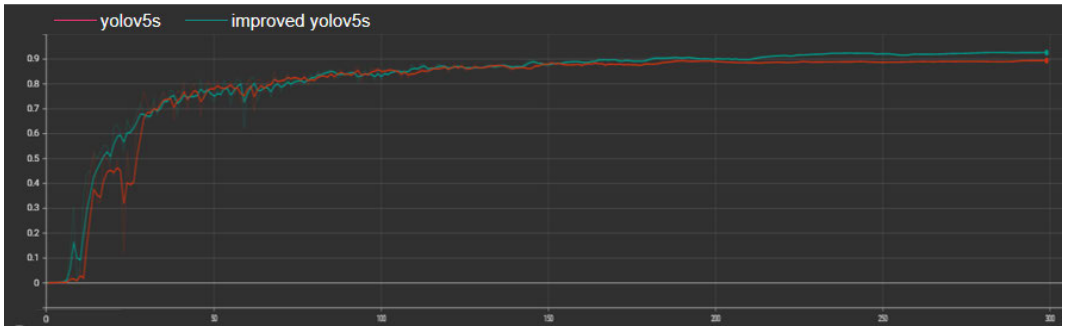

FIGURE 10. Comparison of training results.

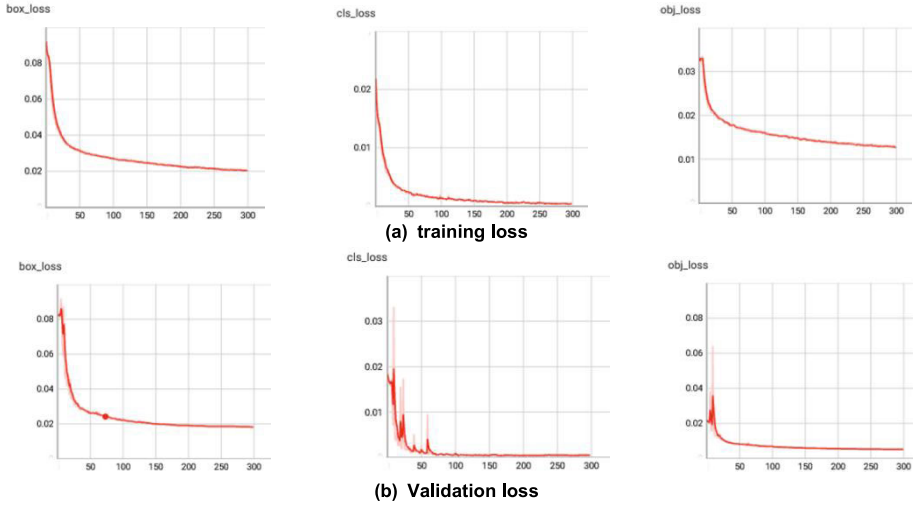

FIGURE 11. Loss curve during the training process.

TABLE 4. Comparison of the results of training set before and after yolov5s improvement.

	Category	AP	mAP@0.5	Model file (MB)
yolov5s	corn	93.2%	89.4%	14.4
	weed	85.6%		
Improved yolov5s	corn	96.3%	92.6%	10.8
	weed	88.9%		

TABLE 5. Evaluation results of test set under different conditions.

	Category	AP	mAP@0.5	Speed
yolov5s	corn	96%	91.7%	6ms/f
	weed	87.4%		
Improved yolov5s	corn	97%	93.4%	3.2ms/f
	weed	89.8%		

TABLE 6. Test results of field precision spraying operation.

Category	Number	Successful detection	TP	FN	FP	Precise spraying	Incorrect spraying
Corns	600	598	556	42	--	--	68
Weeds	465	418	386	32	60	380	126

(a) Original picture

(b) Turn down the brightness

(c) Add noise

(d) Occluded

FIGURE 12. The detection effect pictures under different conditions.

model is improved, indicating that the overall recognition performance of the model is improved. The AP value of corn and weeds was improved. And the model file exported by the improved model training is only 10.8 MB, which is 3.6 MB smaller than the original model. It shows that the improved model maintains high recognition accuracy of corn and weeds, while the model file size is effectively reduced, making the model easier to embed. In conclusion, the improved model in this study demonstrates optimal performance in the detection of corn and weeds and meets the standards for precise identification, while also accomplishing the anticipated improvement goals.

FIGURE 13. Placing the water-sensitive Paper. Divided into three situations, (a) Put water-sensitive paper near corn, (b) Put water-sensitive Paper in a place without corn and weeds, (c) Put water-sensitive paper near weeds.

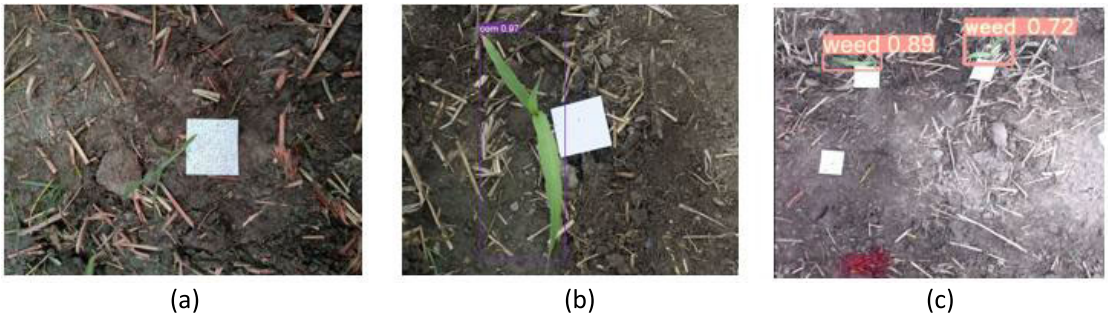

FIGURE 14. Effect Diagram of Precision Spraying Experiment. (a) Precision spraying effect, (b) Recognition effect of corn, (c) Recognition effect of weed.

2) TEST RESULTS ANALYSIS OF MODEL

To further evaluate the detection performance of improved yolov5s, test work was performed on the test set after the training was completed. There are 3800 pictures in the test set, and the photos in the test set were fed into the model for testing. The detection results are shown in Table 5.

From Table 5, it can be seen that the improved yolov5s algorithm has better test results on the test set. The mAP@0.5 value of the model reaches 93.4%, which is 1.7% higher than that of the original model. The AP values for corn and weeds are 97% and 89.8%, respectively. Compared with the original model, the AP value for weeds increased by 2.4%, and the recognition accuracy for corn increased by 1%. The improved yolov5s model takes 3.2 ms to detect each image, which is 2.8 ms less than the original model. As can be observed from the data above, the improved yolov5s model performs better than the original model in both recognition effect and detection speed. Since the images in the test set were all from images in the field, the improved yolov5s model had a breakthrough in corn and weed identification detection applications. It showed high recognition accuracy during the testing process, which laid the foundation for the accurate identification of corn weeds and accurate spraying.

Due to the complex field operation environment, the recognition system is easily affected by the environment such as light, noise, and occlusion when the precision spraying robot is operating in the field, causing the recognition performance of the model to be reduced. Therefore, the images of the test set were processed by using random brightness adjustment, random noise addition, and random occlusion. The detection effect graph is shown in Figure 12. As can be seen from the figure, the improved model can still recognize corn and weeds in more complex scenes, indicating that the model has good generalization performance.

F. RESULTS ANALYSIS OF FIELD PRECISION SPRAYING EXPERIMENT

To evaluate the feasibility of a precision spraying robot, this study carried out precision spraying experiments on ecological unmanned farms. The planting row gap of the corn test field in the unmanned farm is 60cm, and the plant spacing is 25cm. The water-sensitive paper is placed as shown in Figure 13. In the spray application system, the USB camera is placed perpendicular to the ground, about 50cm from the ground, and the camera is placed in front of the nozzle, with the distance between the camera and the nozzle at 10 cm. In the experiment, the speed of the precision spraying robot is 0.2m/s.

The precision spraying robot used a camera to obtain video from between the corn rows and transmitted the video to the target detection unit in real time. The improved yolov5s model detected corn and weeds in the video and obtained the target information in the video, the spraying control unit controlled the opening and closing of the pump according to the target information and finally carries out the precise spraying operation of the weeds. There were 600 corn and 465 weeds during the experiment, and the test results are shown in Table 6, and the Precision spraying effect is shown in Figure 14.

In the table, TP indicates corn detected as corn and weed detected as a weed. FN indicates corn detected as weed or weed detected as corn. FP indicates detected as corn and weed when corn and weed were not present. Precise spraying indicates the number of plants that were accurately identified as weeds and successfully sprayed, and incorrect spraying indicates the number of plants that were sprayed when the target was not a weed.

It can be seen from Table 6 that when the precision spraying robot is working in the field, the accuracy of identifying weeds in the field is 83%, the accuracy of identifying corn is 92%, and the probability of corn being identified as weeds is 7%. The probability of correctly identifying weeds and accurately spraying is 81%. The time to detect a single picture is 30ms. From the above data, it can be seen that the application of precision spraying robots in the field is feasible. The improved yolov5s model has a better recognition effect on corn and weeds in the field operation environment and still has a high recognition rate in the environment of wind, occlusion, and weak light, which proves that the improved yolov5s model is robust and effective.

In precision spraying experiments, there are also cases of mistakenly sprayed, unidentified, mistakenly identified, and weeds that are successfully identified but not sprayed. The reason for mistakenly sprayed and mistakenly identified is that there are many wheat seedlings in the corn field, and the model identified some wheat seedlings as weeds. The reason why weeds are successfully identified but not sprayed is that the precision spraying system did not open the pump in time. The reason why the target is not identified may be that the characteristics of corn and weeds cannot be accurately captured by the model due to the wind and the jolt of the robot, resulting in the inability to identify them. There are still many challenges in precision spraying weeding operations, but this study verifies the feasibility of precision spraying operations and provides data support for precision spraying robots.

IV. CONCLUSION

In this study, a lightweight model based on the improved yolov5s was proposed to detect corn and weeds in complex environments in the field, and a precision spraying robot was designed. At first, the corn and weeds image data were collected, and the dataset was built using data enhancement methods based on category balance and agronomic characteristics. Using feature map visualization experiments, we found the problem of the yolov5s model in the feature extraction stage, which can not emphasize the important feature information and suppress irrelevant feature information. Therefore, the SENet network was introduced in the feature extraction network, which enables the model to focus more on important features and suppress the extraction of unimportant features to improve the recognition performance of the model. In this study, the C3-Ghost-bottleneck module was designed using the Ghost module and replaced the C3 module in the yolov5s structure, thereby reducing the number of parameters and computation, making the yolov5s model more lightweight and easier to deploy at the edge. The results show that the overall performance of the improved yolov5 s model was better than that of the original model. It had better detection accuracy and detection speed. Compared with the original model, the map@0.5 value is improved by 3.2%, and the file size is reduced by 3.6MB. The AP value for corn recognition increases from 93.2% to 96.3%, and the AP value increases by 3.1%. The AP value for weeds increases from 85.6% to 88.9%, an increase of 3.3%. The time to detect a single image is 3.2 ms. The robustness of the model is demonstrated by detecting images with different brightness, noise, and random occlusion. Based on the improved yolov5s model, the improved model was deployed on Jetson AGX Xavier, the precision spraying system was designed and a precision spraying robot was built. The precision spraying experiments of weeds were conducted in the field, the accuracy of identifying weeds in the field is 83%, the accuracy of identifying corn is 92%, and the probability of corn being identified as weeds is 7%. The probability of correctly identifying weeds and accurately spraying is 81%. The time to detect a single picture is 30ms. The results show that precision spraying operation is feasible and the precision spraying robot can achieve precision spraying operation with high accuracy, and it also verifies that the improved yolov5s model has better recognition performance in the field.

In future work research, we intend to further improve the detection ability for corn and weeds. Additionally, we will continue to optimize the precision spraying robot.

REFERENCES

[1] W. Han, X. Peng, L. Zhang, and Y. Niu, "Summer corn yield estimation based on vegetation index derived from multi-temporal UAV remote sensing," *Trans. Chinese Soc. Agricult. Mach.*, vol. 51, pp. 148–155, Jan. 2020.

[2] I. Rajcan and C. J. Swanton, "Understanding maize–weed competition: Resource competition, light quality and the whole plant," *Field Crops Res.*, vol. 71, no. 2, pp. 139–150, Jun. 2001.

[3] V. Partel, S. C. Kakarla, and Y. Ampatzidis, "Development and evaluation of a low-cost and smart technology for precision weed management utilizing artificial intelligence," *Comput. Electron. Agricult.*, vol. 157, pp. 339–350, Feb. 2019.

[4] J. Liu, I. Abbas, and R. S. Noor, "Development of deep learning-based variable rate agrochemical spraying system for targeted weeds control in strawberry crop," *Agronomy*, vol. 11, no. 8, p. 1480, Jul. 2021.

[5] L. Wang, Y. Zhao, S. Liu, Y. Li, S. Chen, and Y. Lan, "Precision detection of dense plums in orchards using the improved YOLOv4 model," *Frontiers Plant Sci.*, vol. 13, Mar. 2022, Art. no. 839269.

[6] K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," *Sensors*, vol. 18, no. 8, p. 2674, 2018.

[7] Y. Chen, Z. Wu, B. Zhao, C. Fan, and S. Shi, "Weed and corn seedling detection in field based on multi feature fusion and support vector machine," *Sensors*, vol. 21, no. 1, p. 212, Dec. 2020.

[8] X. Zhao et al., "Ultrasonic sensing system design andaccurate target identification for targeted spraying," in *Advanced Manufacturing and Automation VII*. Singapore: Springer, 2018, pp. 245–253.

[9] V. Jejčć, T. Godeśa, M. Hoćevar, B. Širok, A. Malnerśić, A. Štancar, M. Leśnik, and D. Stajnko, "Design and testing of an ultrasound system for targeted spraying in orchards," *J. Mech. Eng.*, vols. 7–8, no. 57, pp. 587–598, Aug. 2011.

[10] X. Liu, Y. Li, M. Li, J. Yuan, Q. Fang, and J. Hou, "Design and test of smart-targeting spraying system on boom sprayer," *Trans. Chin. Soc. Agricult. Mech.*, vol. 47, no. 3, pp. 37–44, 2016.

[11] Y. Lan, L. Geng, W. Li, W. Ran, X. Yin, and L. Yi, "Development of a robot with 3D perception for accurate row following in vineyard," *Int. J. Precis. Agricult. Aviation*, vol. 1, no. 1, pp. 14–21, 2018.
[12] H. Huang, J. Deng, Y. Lan, A. Yang, Y. Jiang, G. Suo, and P. Chen, "Automatic difference vegetation index generator for spider mite-infested cotton detection using hyperspectral reflectance," *Int. J. Precis. Agricult. Aviation*, vol. 1, no. 1, pp. 83–88, 2018.
[13] N. Wang, N. Zhang, and F. E. Dowell, "Design of an optical weed sensor using plant spectral characteristics," *Trans. Chin. Soc. Agricult. Mech.*, vol. 44, no. 2, p. 409, 2001.
[14] W. Deng, C.-J. Zhao, X.-K. He, Li-ping Chen, L.-D. Zhang, G.-W. Wu, J. Mueller, and C.-Y. Zhai, "Study on spectral detection of green plant target," *Spectrosc. Spectral Anal.*, vol. 30, no. 8, pp. 2179–2183, 2010.
[15] S. Chen, H. Zou, R. Wu, R. Yan, and H. Mao, "Identification for weedy Rice at seeding stage based on hyper-spectral imaging technique," *Trans. Chin. Soc. Agricult. Mech.*, vol. 444, no. 5, pp. 163–253, 2013.
[16] J.-J.-Q. Armstrong, R. D. Dirks, and K. D. Gibson, "The use of early season multispectral images for weed detection in corn," *Weed Technol.*, vol. 21, no. 4, pp. 857–862, Dec. 2007.
[17] C. J. Gray, D. R. Shaw, P. D. Gerard, and L. M. Bruce, "Utility of multispectral imagery for soybean and weed species differentiation," *Weed Technol.*, vol. 22, no. 4, pp. 713–718, Dec. 2008.
[18] A. Kamilaris and F. X. Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Apr. 2018.
[19] B. Åstrand and A. J. Baerveldt, "An agricultural mobile robot with vision-based perception for mechanical weed control," *Auto. Robots*, vol. 13, no. 1, pp. 21–35, 2002.
[20] N. Li, "Research on machine-vision-based fast crop detection method for robotic weeder," Ph.D. thesis, China Agricult. Univ., Beijing, China, 2017.
[21] W. Lanlan and W. Youxian, "Application of support vector machine for identifying single corn/weed seedling in fields using shape parameters," in *Proc. 2nd Int. Conf. Inf. Sci. Eng.*, Dec. 2010, pp. 1–4.
[22] L. Wu and Y. Wu, "Weed/corn seedling recognition by support vector machine using texture features," *Afr. J. Agricult. Res.*, vol. 4, no. 9, pp. 840–846, 2009.
[23] W. Mao, J. Cao, H. Jiang, Y. Wang, and X. Zhang, "In-field weed detection method based on multi-features," *Trans. Chin. Soc. Agricult. Mech.*, vol. 23, no. 11, pp. 206–209, 2007.
[24] F. Lin, D. Zhang, Y. Huang, X. Wang, and X. Chen, "Detection of corn and weed species by the combination of spectral, shape and textural features," *Sustainability*, vol. 9, no. 8, p. 1335, Aug. 2017.
[25] L. Chen, J.-G. Zhang, H.-F. Su, and W. Guo, "Weed identification method based on probabilistic neural network in the corn seedlings field," in *Proc. Int. Conf. Mach. Learn. Cybern.*, Jul. 2010, pp. 1528–1531.
[26] K. Hu, Z. Wang, G. Coleman, A. Bender, T. Yao, S. Zeng, D. Song, A. Schumann, and M. Walsh, "Deep learning techniques for in-crop weed identification: A review," 2021, *arXiv:2103.14872*.
[27] H. Jiang, C. Zhang, Y. Qiao, Z. Zhang, and W. Zhang, "CNN feature-based graph convolutional network for weed and crop recognition in smart farming," *Comput. Electron. Agricult.*, vol. 174, Jan. 2020, Art. no. 105450.
[28] C.-C. Andrea, B. B. Mauricio Daniel, and J. B. Jose Misael, "Precise weed and maize classification through convolutional neuronal networks," in *Proc. IEEE 2nd Ecuador Tech. Chapters Meeting (ETCM)*, Oct. 2017, pp. 1–6.
[29] H. Pei, Y. Sun, H. Huang, W. Zhang, J. Sheng, and Z. Zhang, "Weed detection in maize fields by UAV images based on crop row preprocessing and improved YOLOv4," *Agriculture*, vol. 12, no. 7, p. 975, Jul. 2022.
[30] L. Cheng, S. Shi-Quan, and G. Wei, "Maize seedling and weed detection based on MobileNetv3-YOLOv4," in *Proc. China Autom. Congr. (CAC)*, Oct. 2021, pp. 5679–5683.
[31] L. Quan, H. Feng, Y. Lv, Q. Wang, C. Zhang, J. Liu, and Z. Yuan, "Maize seedling detection under different growth stages and complex field environments based on an improved faster R–CNN," *Biosystems Eng.*, vol. 184, pp. 1–23, Aug. 2019.

BAOJU WANG was born in 1996. He received the M.S. degree in agricultural engineering from the Shandong University of Technology, Zibo, China, in 2022. He is a member of the National Center for International Collaboration Research on Precision Agricultural Aviation Pesticides Spraying Technology (NPAAC). His research interests include computer vision, image processing, and artificial intelligence.

YU YAN was born in 1997. She received the B.S. degree from the Shandong University of Technology, Zibo, China, in 2021, where she is currently pursuing the M.S. degree in agricultural engineering. Her research interest includes intelligent spraying technology.

YUBIN LAN was born in 1967. He received the B.S. degree in agricultural machinery design and manufacturing from Jilin University, Jilin, China, in 1982, and the Ph.D. degree in agricultural engineering from Texas A&M University, in 1994. He is a Professor with the Shandong University of Technology; a Foreign Academician with the European Academy of Sciences, Arts and Humanities, the Russian Academy of Natural Sciences, and the Georgian National Academy of Sciences; and the Director and the Lead Scientist of the National Center for International Collaboration Research on Precision Agricultural Aviation Pesticides Spraying Technology (NPAAC).

MENG WANG was born in 1999. He is currently pursuing the M.S. degree in agricultural engineering with the Shandong University of Technology. His research interest includes precision agricultural aviation technology and equipment.

ZHIHAO BIAN was born in 1998. He is currently pursuing the M.S. degree in agricultural engineering with the Shandong University of Technology. His research interests include plant protection UAV, mechatronics, and automation control.